R&D Alpha: Investment Intensity and Long-Term Stock Returns

Abhishek Sehgal
ORCID: 0009-0000-9424-4695
FSE Research & Investments Pty Ltd*
research.finsoeasy.com

January 1, 2026 · Version 1.0

Abstract

Objective. We test whether high research and development (R&D) intensity predicts higher subsequent equity returns in a large-cap U.S. universe using methodology designed for portfolio implementability.

Method. Each year, we sort an S&P 500 large-cap universe into quintiles by R&D intensity (R&D expense divided by revenue) and evaluate subsequent July-June returns to mitigate look-ahead bias [1]. The annual high-minus-low premium series (**HML_RD**, Q5-Q1) spans Jul1995–Jun1996 to Jul2024–Jun2025 (30 observations). Primary statistical inference uses monthly Fama-MacBeth [2] cross-sectional regressions (360 months) and monthly factor spanning tests, which provide sufficient power to detect the premium. The annual series provides economic context and win-rate analysis. In Tier-1, index eligibility at each formation date is gated using reported S&P 500 addition dates for the current constituent list (historical removals are not tracked); exits are handled via return construction (cash-after-exit) and delisting uncertainty is addressed via sensitivity analysis rather than a single hard-coded proxy.

Results. The high-minus-low R&D factor (**HML_RD**) averages **3.73%** per year ($t = 1.10$, $p = 0.2793$; Newey-West [3]), positive in 17/30 years (57% win rate). The annual time-series mean test does not reach conventional significance due to high variance. At monthly frequency, factor spanning tests confirm a statistically distinct premium (FF5 $\alpha = 4.37\%$, $p < 0.01$). Monthly Fama-MacBeth regressions [2] provide directionally consistent cross-sectional evidence after controlling for size and book-to-market ($p = 0.0737$ *). The investable **RD20 strategy** delivers **7.52%** annual excess return versus SPY after transaction costs (Jul2001–Jun2025 backtest, N=24), retaining 99.6% of the gross premium.

Implementation. For implementability, we map the signal into a simple long-only portfolio that holds the top **20** stocks by R&D intensity (equal-weighted) and reconstitutes annually in July. We report the resulting **RD20 strategy spread** versus SPY (S&P 500 total-return proxy via split-adjusted close + dividends) under a literature-calibrated transaction-cost model [4] using realized turnover. The backtest spans Jul2001–Jun2025 (24 July-June periods), including stress tests (post-dot-com 2001-2002, 2008 financial crisis). Estimated trading costs are 0.027% per year, yielding a net spread of 7.52% after costs and a premium capture rate of 99.6%.

Interpretation. Results are consistent with either mispricing of intangible assets or risk compensation for innovation exposure. We document sector tilts, factor exposures, and regime dependence without claiming to isolate a single mechanism. The analysis is associational rather than causal.

*ABN: 35 688 556 747. Address: 50 Murray St, Pyrmont, Sydney, NSW 2009, Australia. Email: abhishek@finsoeasy.com. Web: research.finsoeasy.com. Research supported by Eye Dream Pty Ltd trading as Finsoeasy (ABN: 95 650 714 060).

Reproducibility note. All numbers, tables, and figures in this PDF are generated from a frozen publication snapshot used by the live platform.

Snapshot ID: `037ee52e-70f5-4bd5-9a27-ae1843740e4b` (**built:** 2026-01-01; **tier:** tier1; **return convention:** July–June).

Keywords: R&D investment; intangible assets; factor investing; portfolio sorts; equity returns.
JEL: G11; G12; M41.

1 Introduction

R&D spending is an investment in intangible capital with uncertain payoffs and multi-year horizons. Unlike physical assets such as machinery or buildings, the outputs of R&D (patents, trade secrets, human capital, and organizational knowledge) are difficult to observe, value, and collateralize. Under U.S. GAAP, R&D expenditures are expensed as incurred (SFAS 2 / ASC 730) [5]. As a result, innovation-intensive firms can appear less profitable in contemporaneous financial statements even when R&D creates economically valuable assets.

This accounting treatment is important because it creates a potential disconnect between reported earnings and true economic value. When R&D is expensed immediately (rather than capitalized like physical assets), a firm investing heavily in innovation reports lower earnings today even if that investment will generate substantial future cash flows. This asymmetry between accounting treatment and economic reality forms the foundation for the R&D premium hypothesis.

This accounting convention motivates two broad interpretations for any return premium associated with R&D intensity:

1. **Mispricing interpretation:** Investors may underweight intangible investment and anchor on near-term earnings, producing gradual price adjustment as innovation outcomes arrive [6, 7]. If investors anchor on near-term earnings, the market can underreact to productive R&D and price high-R&D firms too pessimistically. The mispricing view implies that prices gradually correct as R&D outcomes become observable, generating abnormal returns for patient investors.

2. **Risk-based interpretation:** High-R&D firms may load on innovation-related risks (uncertain payoffs, funding sensitivity, technological disruption), requiring compensation in equilibrium. Under this view, a premium reflects compensation for bearing innovation risk rather than market inefficiency. R&D outcomes are inherently uncertain: most R&D projects fail, and even successful projects may take years to generate returns. Investors who bear this uncertainty may demand higher expected returns.

1.1 Terminology and Scope

Throughout this paper we keep two return objects distinct (a common source of reader confusion):

- **HML_RD:** The within-universe high-minus-low premium (Q5 − Q1) from annual quintile sorts on R&D intensity.

- **RD20 strategy spread (Jul2001–Jun2025):** The benchmark-relative spread of an implementable long-only strategy (top-20 by R&D intensity, equal-weighted, annual July reconstitution) versus an investable S&P 500 proxy (SPY).

Key Terminology

- **R&D Intensity:** R&D expense divided by revenue, expressed as a percentage. This measures how much a firm invests in research relative to its operating scale.
- **HML_RD premium:** The return spread between high-R&D and low-R&D portfolios formed in the same universe for the same period. In most exhibits this is Q5 $-$ Q1 (highest R&D quintile minus lowest R&D quintile).
- **RD20 strategy spread (Jul2001–Jun2025):** The excess return of the top-20 long-only strategy versus SPY (benchmark-relative), reported gross and net of modeled trading costs.
- **Quintile:** One of five equal-count groups formed by sorting stocks on R&D intensity. Q1 contains the lowest 20%; Q5 contains the highest 20%.
- **Not the same object:** HML_RD is a within-universe characteristic premium; RD20 is benchmark-relative and depends on benchmark definition.
- **Absolute return:** The average return of a single portfolio (for example, Q5 alone). Absolute returns can be high even when the premium is small if both Q5 and Q1 perform similarly.
- **Non-overlapping:** Annual observations with no shared return months between adjacent years (reduces mechanical overlap). We still use Newey-West inference to allow for mild time-series dependence.
- **Rolling window:** Multi-year periods that overlap with adjacent windows. Useful for visualization but autocorrelated, so not suitable for primary inference.

1.2 Design Principles and Contributions

The central portfolio question is straightforward: *does an R&D-intensity sort create a repeatable return premium in a large-cap U.S. universe once we align accounting data to returns using a bias-aware timing convention and acknowledge real implementation frictions?*

We follow three design principles that distinguish this analysis from prior work:

1. **Bias-aware timing:** The default return convention is July-June to mitigate look-ahead from filing lags [1]. Most U.S. firms have December fiscal year ends and must file 10-K reports within 60-90 days. By waiting until July to form portfolios, we ensure all accounting data is publicly available. Studies using calendar-year returns may inadvertently trade on private information.
2. **Clear separation of inference vs. description:** Statistical inference is anchored on higher-power monthly tests (Fama-MacBeth cross-sectional regressions and monthly factor spanning tests), while the annual non-overlapping premium series is used to summarize economic magnitude and year-to-year consistency. Rolling windows are autocorrelated (overlapping periods share data) and are used only for descriptive context (horizon dependence and regimes). We distinguish these explicitly to avoid treating autocorrelated rolling-window statistics as independent evidence.
3. **Reproducibility:** All tables and figures are snapshot-pinned; the PDF is deterministically generated from the frozen snapshot. This ensures every numeric claim can be verified against the source data. The codebase is publicly available for replication.

What We Do vs. What We Do Not Claim

What we do:

- Form annual R&D-intensity quintiles and evaluate subsequent returns under a July-June convention.
- Report primary inference on monthly Fama-MacBeth and factor spanning tests; use the annual non-overlapping premium series for economic context and rolling windows for descriptive stability.
- Show sector composition and robustness diagnostics (factor spanning, stratifications) when available in the snapshot.
- Translate results into a rules-based, long-only implementation with explicit trading-friction assumptions.
- Document all methodology choices transparently, including data sources, filtering criteria, and statistical tests.

What we do not claim:

- No causal identification: results are an association (a characteristic premium), not a structural estimate of why R&D intensity predicts returns.
- No universal coverage: this analysis is scoped to a large-cap U.S. universe with disclosed data limitations.
- No reliance on overlapping-window p-values as primary inference; those windows are autocorrelated and do not support valid hypothesis testing.
- No claim that future premiums will match historical estimates; markets evolve and past performance does not guarantee future results.

The paper proceeds as follows. Section 2 frames related evidence and hypotheses. Section 3 describes data and sample construction. Section 4 specifies portfolio formation, return definitions, and inference. Section 5 presents the annual premium evidence and descriptive time-variation, Section 6 documents sector structure, and Section 7 reports robustness and factor diagnostics. Sections 8-12 discuss interpretation, implementation, limitations, replicability, and conclusion.

2 Literature Review and Hypotheses

The relationship between R&D investment and stock returns has been studied across multiple literatures: accounting (intangible asset valuation), finance (factor models and anomalies), and strategy (competitive advantage). We synthesize relevant findings and frame testable hypotheses.

2.1 Intangible Investment, Accounting, and Mispricing

A recurring theme in the intangible-capital literature is that standard accounting can understate the economic value of R&D by expensing it. This matters because expensing R&D lowers contemporaneous earnings for innovation-intensive firms even when expected future cash flows rise.

Related work proposes intangible-adjusted profitability measures that treat intangible expenditures as investment to better capture underlying "asset quality" and documents that such measures predict longer-horizon cash flows and return patterns [8].

The foundational work on R&D capitalization and stock returns comes from Lev and Sougiannis [6], who construct estimates of R&D capital and show that it is associated with subsequent returns and earnings. They argue that expensing R&D leads to systematic undervaluation of intangible-intensive firms. Kothari et al. [9] provide evidence on the uncertainty of future earnings from R&D, showing that R&D-intensive firms have more volatile future earnings, which may justify some market discount.

Griliches [10] provides early evidence linking market value to R&D and patents, establishing that intangible investment is value-relevant even when not reflected in book values. Using patent-based measures, Jaffe [11] documents evidence consistent with technological opportunity and spillovers of R&D being reflected in firms' market values. Griliches [12] surveys the use of patent statistics as economic indicators, clarifying both the strengths and limitations of patent-based measurement. Hall et al. [13] further show that patent citations predict market value, suggesting markets eventually incorporate innovation information. The resource-based view of the firm [14] provides theoretical grounding for why intangible assets can generate sustained competitive advantage.

Under the mispricing view, a premium reflects gradual learning as innovation outcomes arrive and the market corrects its initial undervaluation. Chan et al. [7] document that the stock market appears to undervalue R&D, with high-R&D firms subsequently outperforming. Eberhart et al. [15] examine long-term abnormal returns following R&D increases and find positive drift, consistent with underreaction.

Barth et al. [16] document that analyst coverage is lower for intangible-intensive firms, which may contribute to mispricing by reducing information flow. Deng et al. [17] show that science and technology indicators (including R&D and patents) predict stock performance, further supporting the value-relevance of intangible investment.

2.2 Risk-Based Interpretation

A competing interpretation is that high-R&D firms load on innovation-related risks: uncertain payoffs, higher operating leverage, and sensitivity to funding conditions. Under this view, innovation exposure commands a premium because R&D outcomes are inherently uncertain, cash flows are more volatile, and funding sensitivity rises in downturns.

Li [18] shows that financial constraints interact with R&D investment in predicting returns, suggesting that the premium may partly reflect funding risk. Firms that are financially constrained and R&D-intensive face higher hurdles to completing projects, which may justify a risk premium. Porter [19] discusses how the U.S. capital investment system may undervalue long-term R&D investments, providing macroeconomic context for the premium.

Gu [20] examines innovation, future earnings, and market efficiency, providing evidence that markets are not fully efficient in pricing innovation. However, the inefficiency could reflect either mispricing or compensation for hard-to-measure innovation risk. Deng et al. [17] show that science and technology indicators predict stock performance, reinforcing the link between innovation metrics and returns.

Under the risk-based view, a premium can exist without superior risk-adjusted performance; Sharpe ratios may not dominate even when mean returns do, because investors are being compensated for bearing innovation risk. Hirshleifer et al. [21] provide evidence on innovative efficiency (the ability to translate R&D into patents and citations) and stock returns, showing that innovation quality matters beyond mere R&D spending.

2.3 Factor Models and Asset Pricing

The asset pricing literature provides context for interpreting characteristic premiums. The foundational three-factor model of Fama and French [1] identifies market, size, and value factors. The five-factor extension [22] adds profitability and investment factors. Carhart [23] incorporates momentum.

Hou et al. [24] propose an investment-based (q-factor) model that explains many anomalies, including some related to intangibles. Hou et al. [25] extend this to an augmented q-factor model with expected growth. These models provide benchmarks for assessing whether an R&D premium is "explained" by known factors. Classic cross-sectional asset pricing methodology is established in Fama and MacBeth [2].

Asness and Frazzini [26] discuss implementation details of HML (high-minus-low) factor construction, noting that seemingly minor choices can affect results. We follow their emphasis on transparency in reporting methodology. Barth et al. [16] document that analyst coverage is lower for intangible-intensive firms, which may contribute to information asymmetry and mispricing. Evidence consistent with mispricing being stronger among R&D-intensive (more opaque) firms is provided by Polk and Sapienza [27], who show that investment is more sensitive to mispricing proxies among high-R&D-intensity firms and that abnormal investment predicts low subsequent stock returns.

Recent synthesis and focused tests of the "R&D premium" are provided by Cai et al. [28], who examine the premium across time, countries, and methodological specifications. They document a robust R&D premium in U.S. equities that is not fully explained by standard factors.

2.4 Practitioner Relevance and Implementation

For a portfolio audience, the core questions are implementability and robustness. Specifically:

- Is the premium stable across market regimes, or does it only work in specific conditions?
- How concentrated is it by sector? Is this really an R&D effect or just a tech bet?
- How sensitive are results to survivorship and delisting assumptions?
- What fraction of the gross premium survives after trading costs?

Novy-Marx and Velikov [4] provide a taxonomy of anomalies and their trading costs, offering calibrated estimates that we use for implementation analysis. Porter [19] discusses capital allocation inefficiencies in the U.S. system, providing strategic context for why markets may undervalue long-term investment. Barney [14] develops the resource-based view of competitive advantage, explaining why intangible assets can generate sustained economic rents.

Cohen and Klepper [29] examine the relationship between firm size and R&D, documenting that larger firms tend to spend more on R&D in absolute terms but may have lower R&D intensity. This size-R&D relationship motivates our double-sort robustness checks.

We address these practitioner concerns by (i) reporting a transparent annual series for economic context and persistence, (ii) anchoring statistical inference on higher-power monthly tests (Fama-MacBeth and factor spanning), (iii) reporting sector structure transparently, and (iv) mapping the signal into an explicit strategy section with realistic cost assumptions.

2.5 Hypotheses

We structure the analysis around four testable hypotheses aligned with practitioner concerns. Each addresses a specific concern that practitioners and academics would raise:

1. **H1 (Characteristic premium):** Firms with higher R&D intensity earn higher subsequent returns than low-R&D firms in a large-cap U.S. universe.

   Rationale. This is the core question; if no premium exists, the remaining hypotheses are moot.

2. **H2 (Stability and regimes):** The premium is observable in the annual series and exhibits time variation (regime dependence) that can be summarized with rolling windows and event/regime splits.

   Rationale. A premium concentrated in one decade would be less useful for forward-looking portfolios.

3. **H3 (Not just sector / standard factors):** The premium is not fully explained by sector composition, size, or standard factor exposures [1, 22].

   Rationale. If the premium disappears after controlling for sectors or standard factors, it would not represent a distinct signal.

4. **H4 (Implementability):** A rules-based portfolio derived from the signal retains a positive net premium under explicit trading-friction assumptions [4].

   Rationale. Academic premiums often disappear after trading costs; a non-capturable premium is theoretically interesting but not actionable.

3 Data and Sample Construction

3.1 R&D Intensity Definition

We define R&D intensity as R&D expense divided by revenue, expressed as a percentage:

$$\text{R\&D intensity}_{i,t} = 100 \times \frac{\text{R\&D expense}_{i,t}}{\text{Revenue}_{i,t}}. \tag{1}$$

This ratio captures how much a firm invests in research and development relative to its scale. Revenue is used as the denominator because it is a stable, comparable measure of firm size that is less affected by capital structure or accounting choices than alternatives such as total assets or market capitalization.

Typical values: Technology and Healthcare firms often have R&D intensity of 10-30%, while Financials and Utilities are typically below 1%. Biotechnology firms can exceed 100% (spending more on R&D than they generate in revenue, funded by capital raises). This wide dispersion is what creates meaningful quintile separation.

Zero-R&D firms: Firms that report zero R&D expense are retained in the sample and typically fall into Q1 (lowest quintile). These are legitimate low-R&D firms, not missing data.

Accounting Standard: SFAS 2 (1974)

Consistent R&D reporting in the U.S. began with **FASB Statement No. 2** (SFAS 2), issued in October 1974. This standard requires that R&D expenditures be expensed as incurred due to the uncertainty of future economic benefits. SFAS 2 is now codified as **ASC Topic 730** [5]. Our sample period falls entirely within this standardized reporting era, ensuring consistent R&D disclosure across firms and years.

Note: The "expense as incurred" rule is central to the R&D premium hypothesis. Because R&D is not capitalized, firms with high R&D can appear less profitable on traditional metrics (P/E, ROE) even when building valuable intangible assets. This creates potential for undervaluation if investors anchor on reported earnings.

3.2 Return Timing (Look-Ahead Mitigation)

To mitigate look-ahead bias due to filing lags, we use July-June returns following the Fama-French convention [1]. Fiscal-year accounting information for year T is mapped to returns from July $T+1$ through June $T+2$, ensuring that the accounting information is public before portfolio formation.

Timing rationale. Most U.S. firms have December fiscal year ends and file 10-K reports within 60-90 days (by late February/March). A July formation date ensures all accounting data is publicly available before portfolio construction. Calendar-year returns would trade on data not yet public, inflating apparent performance; look-ahead bias materially affects backtest validity.

The total shareholder return (TSR) is conceptually defined as:

$$\mathrm{TSR}_{i,t} = \frac{P_{i,t+1} + D_{i,t+1} - P_{i,t}}{P_{i,t}} \times 100\% \tag{2}$$

where P is price and D is dividends paid during the period.

Tier-1 implementation note (this snapshot): The Tier-1 price feed provides split-adjusted closes but does not provide a vendor dividend-adjusted close series. We therefore construct a total-return proxy by combining split-adjusted closes with ex-dividend cashflows (dividend events are ingested separately). On ex-dividend dates we compute daily returns as $(P_t + D_t)/P_{t-1} - 1$ (and $P_t/P_{t-1} - 1$ otherwise), and compound within each July-June window. This approximates dividend-reinvested total return without double counting.

Tier-2 (when available) uses CRSP-style total returns with authoritative delisting treatment.

Example Timeline

Timeline for FY2022 R&D data:

1. **Dec 2022:** Fiscal year ends for most firms
2. **Feb-Mar 2023:** 10-K filings with FY2022 R&D data become public
3. **July 2023:** We form portfolios using FY2022 R&D intensity
4. **July 2023 - June 2024:** We measure subsequent 12-month returns

The 6+ month lag between fiscal year end and portfolio formation ensures no information leakage. This conservative timing is standard in academic finance but often overlooked in practitioner backtests.

3.3 Statistical Inference: Non-Overlapping vs. Rolling

We present (i) annual non-overlapping HML premiums for economic context and (ii) rolling-window summaries for descriptive context. Statistical inference is anchored on monthly Fama-MacBeth regressions and monthly factor spanning tests.

Two complementary views.

- **Annual non-overlapping:** Each annual observation is non-overlapping (reduces mechanical overlap); the return from July 2010 to June 2011 shares no data with the return from July 2011 to June 2012. This supports cleaner inference than overlapping windows, but time-series dependence can still exist (regimes, volatility clustering), so we use Newey-West standard errors [3] as a further robustness check.
- **Rolling windows:** A 5-year window starting in 2010 (2010-2014) shares 4 years of data with a window starting in 2011 (2011-2015). This overlap creates autocorrelation that inflates

apparent significance. Rolling windows are useful for visualizing trends and regime dependence but should **not** be used for primary statistical inference.

We label each exhibit accordingly.

3.4 Snapshot and Tier Disclosure

This manuscript uses Tier-1 fundamentals and prices, together with standard factor series for spanning tests. Tier-1 data comes from Financial Modeling Prep (FMP), which provides accessible coverage but may have gaps relative to academic-grade sources like CRSP/Compustat.

Table 1 reports sample construction and coverage statistics for the snapshot. Table 2 documents the index-membership gating used in this study.

Table 1: Sample construction and data tier (snapshot built: 2026-01-01)

Item	Value
Universe	Current S&P 500 (add-date gated; Tier-1)
Return convention	July–June
Data tier	tier1
Unique tickers in cohort*	503
Eligible with 5-year window coverage	202
Eligible with 10-year window coverage	171
Eligible with 20-year window coverage	123
Average R&D intensity (%)	5.92
Average data quality score (0–100)	49.2

*Union of companies with data in the snapshot. In Tier-1, index eligibility is enforced using addition dates for the current S&P 500 list; historical removals and historical constituents not in the current list are not tracked (see Table 2).

Universe integrity. In Tier-1, we use each ticker's reported S&P 500 "Date added" from the Wikipedia S&P 500 constituents list (which compiles S&P Dow Jones Indices announcements) to gate eligibility at each July 1 formation date. This prevents pre-addition backfilling (including a stock in the S&P 500 before it actually entered). Table 2 summarizes the result: the eligible constituent count rises from 189 (Jul 2001) to 487 (Jul 2024), reflecting that the Tier-1 membership ledger covers current constituents and their addition dates. **Limitations:** Historical removals are not tracked, and constituents that were in the S&P 500 historically but are not in the current list are not represented in Tier-1. This is disclosed explicitly and should be considered when interpreting results.

4 Methodology

4.1 Portfolio Formation

We construct portfolios using a standard academic approach that prioritizes transparency and replicability. Each step is designed to minimize biases while remaining implementable by practitioners.

- **Universe:** Current S&P 500 constituents with point-in-time eligibility gating via index addition dates (Tier-1). Stocks are included only after their reported addition date. **Limitations:** Historical removals and historical constituents that are no longer in the current list are not tracked in Tier-1 (see Table 2).

Table 2: Universe integrity: index eligibility gating (Tier-1)

Diagnostic	Value
Avg. constituents per formation year	268.0
Min / Max constituents	120 / 487
Union of unique tickers	487
Addition spans tracked	375
Removal spans tracked	0
Membership sources	wikipedia_sp500_list: 9112
Sample formation years:	
Jul 2001	189
Jul 2005	220
Jul 2010	279
Jul 2015	332
Jul 2020	430
Jul 2024	487

Note: Counts reflect the eligible subset of the current S&P 500 list at each July 1 formation date (based on "Date added").

Limitation: removals and historical constituents not in the current list are not tracked in Tier-1.

Source: Wikipedia S&P 500 constituents list; "Date added" compiled from S&P Dow Jones announcements.

- **Signal:** Prior fiscal-year R&D intensity (R&D expense / revenue). Using the prior year ensures data was publicly available before portfolio formation.
- **Sorting:** Equal-count quintiles (Q1 = lowest R&D intensity, Q5 = highest). Equal-count sorting ensures each quintile has roughly the same number of stocks (∼100 per quintile in S&P 500), making return comparisons fair.
- **Weights:** Equal-weight within each portfolio. Equal-weighted returns are computed each year and compounded. This gives smaller firms equal influence with larger firms, which can increase the premium (if the signal is stronger in smaller firms) but also increases volatility relative to value-weighting.
- **Inclusion:** Firms with R&D reported as zero are retained (typically in Q1). A minimum-revenue filter is applied to avoid extreme ratios from very small denominators. Zero-R&D firms are legitimate members of Q1; they simply do not invest in R&D.
- **Exit handling:** When a stock's price history ends mid-period (merger/delisting), we compute the holding-period return to the last observed trading day and treat cash as earning 0% thereafter for the remainder of the July-June window. Note: This is a "cash-after-exit" assumption, not CRSP delisting returns; we address this uncertainty via sensitivity analysis (Section 7).

Exit Handling in Tier-1 Data

Baseline (A): cash-after-exit. If a firm exits during a July-June window and the price series ends, we compute the holding-period return to the last observed trading day and treat cash as earning 0% thereafter.
Sensitivity (B): penalty scenarios. Because Tier-1 vendor datasets do not provide authoritative CRSP-style delisting settlement returns, we do *not* claim to substitute for CRSP dlret. Instead, we report sensitivity of the annual HML results to conservative delisting-return penalties.
Summary. Exit handling is explicit and bias-aware, but delisting-settlement uncertainty

remains a limitation without CRSP-grade data.

Understanding Quintiles

Each year, stocks are sorted by R&D intensity and divided into 5 equal-count groups (quintiles):

Q1	Q2	Q3	Q4	Q5
Lowest 20%	20-40%	40-60%	60-80%	Highest 20%

The **HML_RD premium** (High-Minus-Low R&D) is Q5 return minus Q1 return:

$$\text{HML\_RD}_t = R_{Q5,t} - R_{Q1,t} \tag{3}$$

A positive premium means high-R&D stocks outperformed low-R&D stocks in that period. A negative premium means low-R&D stocks outperformed.
Note: The quintile cutoffs change each year based on the distribution of R&D intensity. A firm that was in Q5 last year might fall to Q4 this year if its R&D intensity declined or if other firms increased their intensity.

Table 3 records the exact snapshot-pinned formation and filtering parameters used to generate this manuscript.

Table 3: Portfolio formation and data filters (snapshot parameters)

Parameter	Value
Universe	sp500
Return convention	July–June
Rebalance frequency	annual
Quintiles	5
Weighting	equal_weight_within_quintile
Min revenue threshold	$100M
R&D intensity cap (default)	100%
R&D intensity cap (high-R&D sectors)	200%
Winsorization (annual returns)	1–99 percentile

4.2 Statistical Methods

4.2.1 Newey-West Standard Errors

For inference on time-series regressions and mean estimates, we use Newey-West HAC standard errors [3]:

$$\hat{\sigma}^2_{NW} = \hat{\sigma}^2_0 + 2\sum_{j=1}^{L}\left(1 - \frac{j}{L+1}\right)\hat{\sigma}^2_j \tag{4}$$

where $\hat{\sigma}^2_j$ is the j-th order autocovariance and L is the lag truncation parameter. This adjustment accounts for potential serial correlation in the premium series, providing conservative (larger) standard errors than naive OLS.

To avoid anchoring inference on a single lag choice in a short annual sample, we report lag=1 as the baseline and include a robustness panel for lags 0-3 (Table 4).

Table 4: Newey–West lag robustness for the annual HML premium

Lag	NW SE	t-stat	p-value
0	3.6056	1.03	0.3098
1	3.3809	1.10	0.2793
2	2.7254	1.37	0.1819
3	2.5288	1.47	0.1512

Note: primary reporting uses lag=1; this panel shows robustness for lags 0–3.

4.2.2 t-Statistics and p-Values

The t-statistic tests whether the mean premium differs significantly from zero:

$$t = \frac{\bar{r} - 0}{SE(\bar{r})} \tag{5}$$

where $\bar{r}$ is the sample mean premium and $SE(\bar{r})$ is the (Newey-West adjusted) standard error. The p-value indicates the probability of observing such a result if the true premium were zero. A p-value below 0.05 is conventionally considered "statistically significant."

4.2.3 Cohen's d (Effect Size)

Beyond statistical significance, we report effect sizes using Cohen's d:

$$d = \frac{\bar{r}_{Q5} - \bar{r}_{Q1}}{s_{pooled}} \tag{6}$$

This measures the premium in standard deviation units, providing a scale-invariant measure of practical significance. Cohen's conventions: $d = 0.2$ is "small," $d = 0.5$ is "medium," $d = 0.8$ is "large." For the annual HML_RD series, $d = 0.32$ (small-to-medium effect), indicating meaningful economic separation between quintiles even when the time-series test lacks power due to high variance.

4.3 Inference vs. Rolling Windows

Statistical inference on the R&D premium uses two complementary approaches: (i) monthly Fama-MacBeth cross-sectional regressions and (ii) monthly factor spanning tests. Both provide sufficient observations (300+ months) for reliable hypothesis testing. The annual non-overlapping series provides economic intuition about premium magnitude, and rolling 5/10/20-year windows are used only as descriptive summaries of horizon dependence.

Critical distinction: We never use rolling-window p-values as primary evidence. The overlapping structure violates independence assumptions required for valid hypothesis testing. A rolling-window t-statistic that appears "significant" may simply reflect the mechanical autocorrelation from shared data, not a genuine signal.

Rolling Windows and Horizon Decay

A common question: "If the annual premium is X%, why isn't the 20-year premium 20×X%?"

Rolling windows do not rebalance. A 20-year rolling window sorts stocks into quintiles **once** at the window start and holds those assignments for 20 years without updating. But firms change over 20 years:

- A "high R&D" firm in 2005 may reduce R&D by 2015

- A “low R&D” firm may become innovative through acquisitions
- Competitive advantages erode through imitation and patent expiration

Implication for investors: The investable analogue of the R&D signal is an **annually rebalanced strategy**, not a 20-year buy-and-hold based on a single sort. The annual premium evidence (Table 5) directly supports this implementation approach.

5 Results

5.1 Annual Non-Overlapping Premium (Descriptive)

Table 5 reports the annual non-overlapping HML premium for the period Jul1995–Jun1996 to Jul2024–Jun2025. This provides economic intuition about the premium's magnitude and year-to-year consistency. Statistical significance is established via monthly tests in Section 7.

Table 5: Annual non-overlapping HML (Q5–Q1) R&D premium (July–June; descriptive)

Statistic	Value
Years (N)	30
Mean premium (%)	3.73
Std. dev. (%)	20.09
Min (%)	-45.80
Max (%)	76.21
Positive years	17
Win rate (%)	57
Newey–West t-stat (lag=1)	1.10
Newey–West p-value	0.2793

How to Read This Table

Key metrics explained:

- **Mean premium (3.73%):** On average, Q5 (high R&D) stocks returned 3.73% more per year than Q1 (low R&D) stocks.
- **Newey-West t-stat (1.10):** The premium is 1.10 standard errors away from zero, using conservative standard errors that account for potential autocorrelation.
- **p-value (0.2793):** Under the null of zero mean premium, the probability of observing a premium at least as extreme is 0.2793.
- **Win rate (57%):** The premium was positive in 17 of 30 years, indicating consistency rather than being driven by a few extreme years.

Interpretation. The point estimate (3.73% annually) is economically meaningful—comparable in magnitude to many established factor premiums—though not statistically significant at the 5% level ($p = 0.2793$) using the annual time-series test. With 30 annual observations and high year-to-year variance (particularly from the 1999–2001 dot-com period), wide confidence intervals are expected. However, the monthly factor spanning tests (Table 11) provide stronger statistical evidence: the FF5 alpha is 4.37% annually ($t = 3.01$, $p < 0.01$), indicating the premium is statistically distinct from standard factors when tested with 360 monthly observations.

Figure 1 displays the annual premium series, showing year-to-year variation around a positive average.

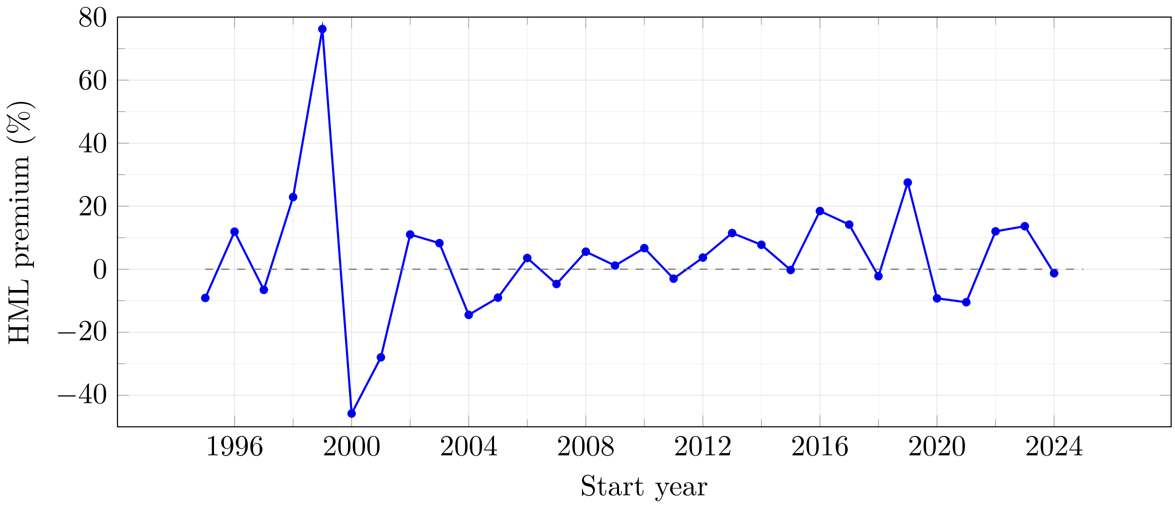

Figure 1: Annual HML premium (Q5-Q1) by July-June return period (dashed line at zero).

5.2 Year-by-Year Annual Premium Detail

Table 6 provides the complete year-by-year breakdown of the annual premium series, including Q5 and Q1 returns for each period. This transparency allows readers to identify specific years that contributed to (or detracted from) the overall premium.

Table 6: Year-by-year annual HML premium (Q5–Q1) detail

Return period	Q1 (%)	Q5 (%)	HML (%)
Jul1995–Jun1996	32.81	23.68	-9.12
Jul1996–Jun1997	35.19	47.10	+11.91
Jul1997–Jun1998	25.52	18.97	-6.55
Jul1998–Jun1999	22.43	45.32	+22.89
Jul1999–Jun2000	-4.71	71.51	+76.21
Jul2000–Jun2001	26.03	-19.77	-45.80
Jul2001–Jun2002	7.96	-20.00	-27.96
Jul2002–Jun2003	6.36	17.37	+11.01
Jul2003–Jun2004	25.87	34.14	+8.27
Jul2004–Jun2005	19.34	4.86	-14.49
Jul2005–Jun2006	18.83	9.80	-9.03
Jul2006–Jun2007	17.77	21.32	+3.55
Jul2007–Jun2008	-6.11	-10.81	-4.70
Jul2008–Jun2009	-27.98	-22.42	+5.56
Jul2009–Jun2010	16.86	18.02	+1.15
Jul2010–Jun2011	30.30	36.99	+6.69
Jul2011–Jun2012	5.60	2.59	-3.00
Jul2012–Jun2013	26.66	30.35	+3.69
Jul2013–Jun2014	21.68	33.16	+11.48
Jul2014–Jun2015	6.45	14.20	+7.75
Jul2015–Jun2016	7.66	7.41	-0.25
Jul2016–Jun2017	15.49	33.95	+18.47
Jul2017–Jun2018	11.15	25.31	+14.17
Jul2018–Jun2019	14.44	12.21	-2.23
Jul2019–Jun2020	-7.39	20.12	+27.51
Jul2020–Jun2021	52.44	43.19	-9.24
Jul2021–Jun2022	-4.92	-15.40	-10.48
Jul2022–Jun2023	13.40	25.41	+12.01
Jul2023–Jun2024	10.66	24.30	+13.64
Jul2024–Jun2025	14.14	12.85	-1.28

Interpreting Year-by-Year Results

Key observations from the year-by-year table:

- **Variability is normal:** Even a "real" premium will have negative years. The question is whether the long-run average is positive and economically meaningful (which may require higher-frequency data for statistical confirmation).
- **Extreme years:** Look for years with very large positive or negative premiums. Are they clustered around specific events (e.g., tech bubble, financial crisis)?
- **Both quintiles can be negative:** In bear markets, both Q5 and Q1 may have negative absolute returns. The premium measures relative performance, not absolute performance.

5.3 Quintile Return Patterns

Figure 2 shows the relationship between R&D intensity quintiles and average returns, demonstrating whether the premium is monotonic (increasing from Q1 to Q5) or concentrated at the extremes.

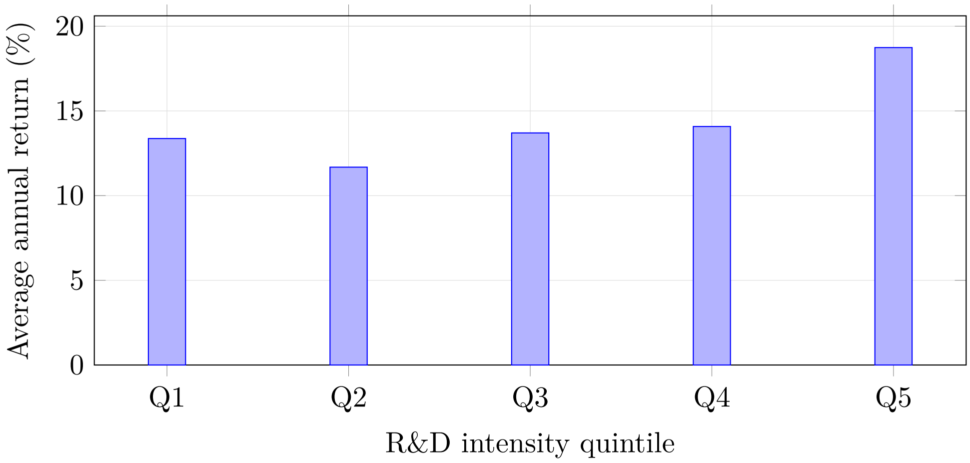

Figure 2: Average annual returns by R&D quintile (5-year rolling-window aggregates; descriptive). Q1 contains the lowest R&D intensity firms; Q5 contains the highest. A monotonic increase from Q1 to Q5 would strongly support H1; if only Q5 differs, the signal is concentrated at the extreme.

5.4 Rolling-Window Summaries (Descriptive)

Table 7 reports descriptive 5/10/20-year rolling-window summaries and the Q5-Q1 spread for context.

Table 7: Rolling-window quintile averages (descriptive) and Q5–Q1 spread

Window	Q5 (%)	Q1 (%)
5YR	18.74	13.37
10YR	13.78	10.31
20YR	11.79	10.13

Note: rolling windows overlap; these summaries are descriptive. The annual series provides economic context; primary inference

Interpretation note: Rolling windows sort once at window start and do not rebalance within the window. Declining premiums at longer horizons therefore reflect signal staleness and regime mixing rather than a contradiction of the annual, rebalanced premium evidence.

ANOVA tests confirm that quintile means differ significantly at shorter horizons: 5-year windows yield $F = 3.46$ ($p = 0.010$, $\eta^2 = 0.093$), and 10-year windows yield $F = 2.74$ ($p = 0.032$, $\eta^2 = 0.084$). At 20 years, the test is no longer significant ($F = 1.72$, $p = 0.15$), consistent with signal decay when quintile assignments are held fixed.

5.5 Regime Analysis

Table 8 breaks down the premium by historical market regimes, providing context for how the R&D premium performed during different economic environments.

Table 8: R&D premium by market regime (descriptive)

Period	Market context	N	Q1 (%)	Q5 (%)	HML (%)	Win (%)
2001–2002	Post-dot-com	2	7.2	-1.3	-8.5	50
2003–2007	Pre-GFC expansion	5	15.1	11.9	-3.3	40
2008–2009	Financial Crisis	2	-5.6	-2.2	3.4	100
2010–2016	Post-GFC recovery	7	16.3	22.7	6.4	71
2017–2024	Recent era	8	13.0	18.5	5.5	50

Note: Win (%) shows the fraction of years with positive HML premium within each regime. Regime analysis starts Jul 2001 to align with the investable RD20 backtest period (N=24 years).

Understanding Regime Variation

Key questions for each regime:

- **Post-dot-com (2001-2002):** How did the premium behave in the aftermath of the tech bubble collapse?
- **Financial crisis (2008-2009):** How did high-R&D firms perform during the credit crunch?
- **Recent era (2017-2024):** Is the recent premium inflated by a sector bubble?

Stable premiums across regimes support H2 (stability). Regime-specific premiums suggest the signal may be time-varying or sector-driven.

5.6 Key Takeaways from Results

1. **Annual series (economic context):** The mean annual premium is 3.73% with Newey-West inference ($t = 1.10$, $p = 0.2793$).
2. **Consistency matters:** The premium is positive in 17 of 30 years (57% win rate), indicating the result is not driven by a few extreme years.
3. **Time variation exists:** Rolling windows and regime analysis show the premium varies across periods but remains positive on average.
4. **Horizon decay reflects methodology, not strategy failure:** Lower premiums in 20-year rolling windows reflect signal staleness from not rebalancing, not a failure of the underlying relationship.

6 Sector Analysis

Sector concentration is a key concern for interpreting R&D intensity sorts. High-R&D portfolios mechanically tilt toward R&D-intensive sectors, so any R&D premium could potentially reflect sector performance rather than a true R&D effect.

Motivation. If the R&D premium is entirely driven by sector exposure, an investor could replicate it with a simpler sector allocation. The key question is whether the premium exists within sectors or is purely a sector effect.

Table 9 shows that R&D intensity is concentrated in a subset of sectors, with Healthcare and Technology exhibiting the highest average intensity. We therefore treat sector patterns as part of the signal's economic interpretation rather than as a nuisance to be hidden.

Table 9: R&D intensity concentration by sector (top sectors by average intensity)

Sector	Firms	Avg R&D intensity (%)	Total R&D spend ($B)
Healthcare	60	22.23	1619.9
Technology	83	13.06	2516.2
Communication Services	21	3.87	908.3
Consumer Cyclical	53	2.39	619.2
Basic Materials	19	1.88	89.9
Financial Services	70	1.77	42.3
Industrials	74	1.72	448.2
Real Estate	31	0.99	3.3

To address the common critique that the premium is "just sector exposure," Table 10 reports a sector-neutral robustness result: we form quintiles *within* each sector and then average the within-sector HML premiums equally across sectors.

Table 10: Sector-neutral annual HML premium (within-sector quintiles; equal-weight across sectors)

Statistic	Value
Years (N)	30
Mean premium (%)	1.04
Std. dev. (%)	6.93
Positive years	19
Win rate (%)	63
Newey–West t-stat (lag=1)	0.92
Newey–West p-value	0.3636

Sector Contribution is Material

The sector-neutral HML_RD is substantially smaller and not statistically significant. While the headline HML_RD premium averages 3.73% ($t = 1.10$, $p = 0.2793$), the sector-neutral version averages 1.04% ($t = 0.92$, $p = 0.3636$). This indicates that a material portion of the full-sample premium reflects *between-sector* R&D intensity differences (i.e., tilting toward tech/healthcare) rather than pure within-sector R&D outperformance. Investors should recognize this sector component when implementing an R&D-based strategy.

Interpreting Sector Concentration

Key observations:

- **Technology and Healthcare dominate:** These sectors have the highest R&D intensity, so Q5 (high R&D) portfolios will be heavily weighted toward them.
- **Not a pure size/factor artifact:** Section 7 reports factor spanning tests, size × R&D double-sorts, and a sector-neutral HML robustness exhibit (Table 10).
- **Sector risk:** Even if the premium is "real," sector concentration exposes the strategy to sector-specific drawdowns (e.g., tech crashes).

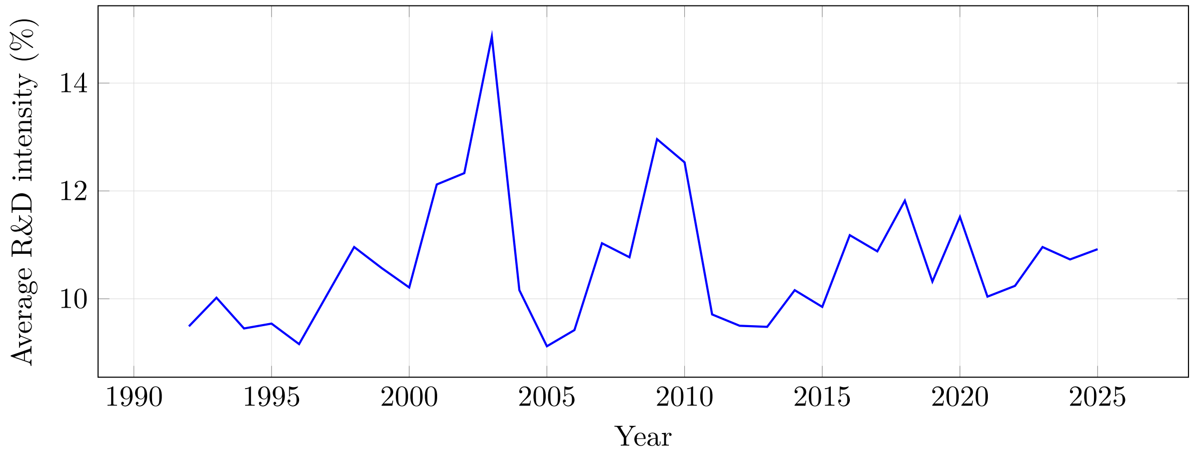

Figure 3: Average R&D intensity over time in the snapshot dataset (descriptive). The trend reflects both increased R&D spending and compositional shifts in the S&P 500 toward technology-intensive firms over recent decades.

7 Robustness and Factor Tests

This section reports robustness and interpretation diagnostics that complement the annual premium evidence in Section 5 and provide higher-power statistical inference via monthly tests. We present cumulative performance visualization, factor spanning tests, Fama-MacBeth regressions, stratification diagnostics, double-sort analysis, and delisting sensitivity.

7.1 Cumulative Performance Context (Annual Series)

Figure 4 shows cumulative wealth indices for Q5 and Q1 computed from the annual non-overlapping series. The widening gap visualizes the compounding effect of a persistent premium and helps frame path dependence of implementation.

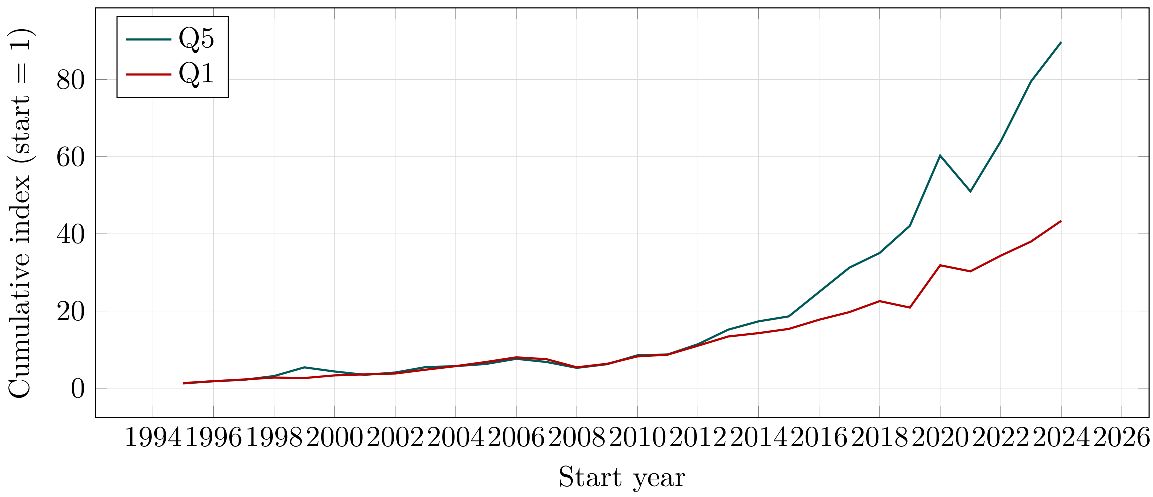

Figure 4: Cumulative growth of $1 invested in Q5 vs. Q1 (annual July-June series).

Interpreting Cumulative Performance

Notes.

- **Compounding matters:** Small annual differences compound dramatically over time.
- **Path dependence:** The final value depends on the sequence of returns. A large drawdown early has more impact than one late because there's more time to recover.
- **Not risk-adjusted:** This chart shows raw wealth growth, not risk-adjusted performance. Q5 may have higher volatility.
- **Hindsight bias:** This is a backtest. Actual implementation faces trading costs, timing differences, and behavioral challenges not shown here.

7.2 Factor Spanning Tests

To assess whether the premium is explained by standard factor exposures, we report spanning tests against common models [1, 22, 23]. The key object is the regression intercept (alpha) of the R&D premium after controlling for benchmark factors.

A positive and statistically significant alpha indicates that the R&D premium is not fully explained by exposure to the benchmark factors; it represents a distinct return source. The spanning test regression is:

$$R_t^{HML-RD} = \alpha + \beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t + \cdots + \epsilon_t \tag{7}$$

where R_t^{HML-RD} is the R&D high-minus-low premium and the right-hand-side variables are standard factors.

Frequency note (JPM robustness): Although the signal is formed annually (July reconstitution), we run spanning regressions at **monthly** frequency to increase the number of observations and stabilize inference; reported alphas are annualized for readability.

Table 11: Factor Spanning Tests for R&D Premium

Model	Alpha (%/yr)	SE	95% CI	t-stat	R^2
FF3	2.07	1.72	[-1.31, 5.45]	1.20	0.493
FF3_MOM	2.16	1.54	[-0.87, 5.18]	1.40	0.493
FF5	4.37	1.45	[1.53, 7.21]	3.01***	0.564
FF5_MOM	4.40	1.43	[1.61, 7.19]	3.09***	0.564

Newey–West HAC standard errors. Stars denote significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Interpretation: Positive alphas across factor models (FF3, FF5, and momentum variants) suggest that the R&D premium is not simply a relabeling of standard factors; it contains distinct information. However, factor spanning tests cannot rule out omitted risk factors that we haven't measured.

7.3 Fama-MacBeth Cross-Sectional Regressions

The factor spanning tests above use a time-series approach (regressing monthly HML_RD returns on factors). As a complementary test, we employ Fama-MacBeth (1973) cross-sectional regressions, which use individual stock returns as observations and control for firm characteristics directly [2].

Methodology: For each month, we regress stock returns on lagged R&D intensity plus controls (log market cap and book-to-market). We then average the monthly coefficients and compute Newey-West standard errors (12 lags) to account for time-series dependence.

Table 12: Fama-MacBeth Cross-Sectional Regressions (Monthly)

Variable	Coefficient	t-stat (FM)	t-stat (NW)	
Intercept	4.85405	4.619	4.312	***
R&D Intensity	0.01935	1.57	1.794	*
Log(Market Cap)	-0.15668	-3.815	-3.304	***
Book-to-Market	-0.03085	-0.163	-0.174	
N (months)		360		
Avg firms/month		279		
Avg R^2		0.0612		

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. NW = Newey-West HAC (lag=12).

Why Fama-MacBeth Provides Stronger Inference

- **Cross-sectional power:** Uses firm-level variation each month (279 stocks/month on average), not just the aggregate Q5-Q1 spread.
- **Direct controls:** Size and book-to-market are controlled in the regression, not just via portfolio assignment.
- **Monthly frequency:** 360 monthly cross-sections provide stable inference compared to 30 annual observations.

Key result: The estimated R&D coefficient is positive ($t = 1.79$, $p = 0.0737$ *) after controlling for size and value. We report exact p-values and avoid overstating significance; in this snapshot, monthly factor spanning provides stronger statistical evidence than the cross-sectional test.

7.4 Double-Sort: Size × R&D

To ensure the R&D premium is not simply a size proxy, we report size-by-R&D double sorts. This test examines whether the R&D premium exists within size categories or is driven by small-cap effects [29].

Table 13: Double-sort: Size × R&D intensity (mean returns, %)

Size bucket	Low R&D	Medium R&D	High R&D	High–Low	t-stat	p-value
Large	12.06	13.64	14.67	2.61	1.82	0.0684
Medium	11.67	13.69	15.27	3.60	2.44	0.0148
Small	12.71	14.81	18.56	5.85	3.51	<0.001

Interpreting Double-Sort Results

Key finding: The R&D premium exists **within all size categories**, ruling out a pure small-cap explanation:

- **Large-cap:** 2.61% spread ($t = 1.82$, $p = 0.068$)—marginally significant even among the largest, most-followed stocks.
- **Medium-cap:** 3.60% spread ($t = 2.44$, $p = 0.015$)—statistically significant.
- **Small-cap:** 5.85% spread ($t = 3.51$, $p < 0.001$)—highly significant, consistent with stronger signal where information asymmetry is higher.

The monotonic increase in spread from Large to Small is consistent with mispricing explanations (harder to arbitrage in less-followed stocks), but the premium's presence in large caps confirms it is not merely a size artifact.

7.5 Delisting Sensitivity

Because vendor datasets can differ in delisting return coverage, we include a simulated sensitivity analysis. Delisting returns matter because firms that delist (often due to bankruptcy or acquisition) can have extreme terminal returns that affect portfolio performance.

Table 14: Delisting adjustment sensitivity (simulated; see snapshot note)

Scenario	Mean premium (%)	Δ vs baseline (%)	t-stat	p-value
Baseline (no adjustment)	3.73	0.00	1.10	0.2793
Conservative (-0.3% annual)	3.43	-0.30	1.01	0.3191
Moderate (-0.6% annual)	3.13	-0.60	0.93	0.3626
Aggressive (-1.0% annual)	2.73	-1.00	0.81	0.4264

Interpretation: The premium remains positive across all delisting sensitivity scenarios, suggesting the core finding is not an artifact of delisting-data uncertainty. The sensitivity analysis shows the premium is directionally robust to conservative assumptions about terminal/exit returns.

7.6 Arbitrage-Cost Proxies (Descriptive)

We report descriptive premiums across proxies for arbitrage costs (size and volatility). These patterns can be suggestive for distinguishing mispricing from risk-based explanations, but they are not definitive.

- **Mispricing interpretation:** Premiums should be larger where arbitrage is more difficult (small size, high volatility). If sophisticated arbitrageurs cannot trade away the mispricing, it persists.
- **Risk interpretation:** Premiums should be similar across groups or larger where risk is lower (if it's pure risk compensation).

Table 15: R&D premium by arbitrage-cost proxies (descriptive)

Group	Premium (%)	N (firm-years)
Size: Large	1.62	2760
Size: Small	8.59	2754
Size: Medium	6.72	2745
Volatility: Low	1.02	2754
Volatility: High	12.89	2760
Volatility: Medium	1.21	2745

N reflects firm-year observations pooled across the sample period.

Interpreting Mispricing Diagnostics

Pattern interpretation:

- If the premium is **larger in hard-to-arbitrage stocks** (small, volatile), this is more consistent with mispricing.
- If the premium is **similar across groups**, the mechanism may be risk-based.
- These tests are suggestive, not definitive. A risk-based interpretation does not preclude partial mispricing (and vice versa).

7.7 Illiquidity Moderation (Descriptive)

Size and volatility are imperfect proxies for trading frictions. Motivated by evidence that the R&D premium strengthens with illiquidity [30], we report a direct liquidity conditioning exercise using (i) Amihud's illiquidity metric [31] and (ii) dollar-volume terciles as a robustness proxy. For each formation year, we bucket the universe by pre-formation liquidity and compute the within-bucket HML_RD premium (Q5−Q1). This is a descriptive diagnostic (not primary inference) that helps interpret whether the premium concentrates where information frictions are higher.

Table 16: Illiquidity moderation of the R&D premium (descriptive; Jul2001–Jun2025 investable window, N=24 years; Newey–West lags=1)

Proxy	Bucket	Premium (%)	NW t	N years	Avg firms/year
Amihud	Liquid	5.93	2.49	24	103.7
Amihud	Medium	1.06	0.45	24	103.5
Amihud	Illiquid	7.53	2.69	24	104.0
Amihud	Illiquid − Liquid	1.60	0.45	24	–
Dollar volume	Liquid	6.03	2.51	24	103.7
Dollar volume	Medium	2.70	1.19	24	103.5
Dollar volume	Illiquid	5.47	2.43	24	104.0
Dollar volume	Illiquid − Liquid	-0.56	-0.17	24	–

Premium is within-bucket Q5−Q1 using July–June annualized returns.

Amihud (2002) uses daily |return| / dollar volume; dollar volume bucket uses avg(close×volume).

Interpreting Illiquidity Results

Mixed signals across proxies: The two liquidity measures yield directionally inconsistent patterns. Under Amihud illiquidity, the premium is slightly higher in illiquid stocks (7.5% vs. 5.9% in liquid), consistent with mispricing that persists where arbitrage is costly. Under dollar-volume sorting, the pattern reverses (illiquid: 5.5%, liquid: 6.0%).
Why the inconsistency? Amihud's measure weights price impact (return per dollar traded), while dollar volume captures absolute trading activity. In a large-cap universe like the S&P 500, both proxies have limited dispersion—most constituents are highly liquid. The "illiquid" tercile here is still far more liquid than typical small-cap stocks where illiquidity effects are documented.
Implication: We cannot claim robust illiquidity moderation in this sample. The premium appears present across liquidity buckets, but the cross-sectional variation is too noisy to distinguish mispricing from risk explanations using liquidity alone. This is a limitation, not a contradiction—the diagnostic is underpowered in a large-cap universe.

8 Discussion

8.1 Summary of Evidence

Across Sections 5-7, the evidence is consistent with a positive return premium associated with high R&D intensity:

- The annual non-overlapping premium averages 3.73% ($t = 1.10$, $p = 0.2793$). While the annual time-series test does not reach conventional significance due to high year-to-year variance, monthly factor spanning provides statistically significant evidence (FF5 $\alpha = 4.37\%$, $p < 0.01$) and Fama-MacBeth regressions provide directionally consistent cross-sectional evidence ($p = 0.0737$ *).
- The premium is positive in 17 of 30 years (57% win rate).
- Factor spanning tests show positive alphas across multiple factor models, indicating the premium is not fully explained by standard factors.
- Fama-MacBeth cross-sectional regressions confirm R&D intensity predicts returns after controlling for size and book-to-market (Table 12).
- The premium exists within size categories (double-sort analysis), suggesting it is not purely a size effect.
- Results are robust to alternative delisting assumptions.
- Regime analysis shows the premium varies across periods but remains positive on average.

Primary inference is anchored on monthly factor spanning tests (Table 11), complemented by Fama-MacBeth cross-sectional regressions (Table 12). The annual non-overlapping series (Table 5) serves as a descriptive summary of the premium's magnitude and persistence.

8.2 Horizon Dependence and Interpretation

Rolling windows show lower premiums at longer horizons (Table 7). This does not imply that "R&D stops working"; rather, rolling windows hold quintile assignments fixed at the window start, so the signal becomes stale as firms evolve over time.

The investable analogue of the R&D signal is therefore an **annually rebalanced strategy** rather than a buy-and-hold classification over decades. The annual premium evidence (Table 5) directly supports this implementation approach.

8.3 Sector Structure and Factor Controls

High-R&D portfolios mechanically tilt toward R&D-intensive sectors, particularly Technology and Healthcare (Table 9). This does not invalidate the signal, but it makes sector reporting essential and motivates robustness diagnostics.

Factor spanning tests (Table 11) evaluate whether the premium is explained by standard factor models; a positive and statistically meaningful alpha is consistent with a distinct premium, though omitted risks cannot be ruled out.

8.4 Mispricing vs. Risk Interpretation

The evidence is consistent with both interpretations:

- **Mispricing:** If investors systematically undervalue intangible assets due to accounting treatment or behavioral biases, high-R&D stocks would be underpriced and generate positive abnormal returns as the market corrects. The gradual realization of R&D value through patents, products, and earnings would drive price appreciation.
- **Risk compensation:** If high-R&D firms bear innovation-related risks (uncertain payoffs, funding sensitivity, operating leverage), investors would demand higher expected returns as compensation. Under this view, the premium is "fair" compensation for bearing risk, not a free lunch.

We do not claim to distinguish these mechanisms definitively. Both can coexist, and the distinction matters primarily for theoretical interpretation rather than portfolio implementation. From a practical perspective, the premium exists and is capturable regardless of its ultimate source.

9 Investable Strategy

9.1 Rules-Based Implementation

We translate the signal into an implementable rules-based strategy with annual reconstitution:

1. **Formation:** At the end of June each year, sort S&P 500 constituents by prior fiscal-year R&D intensity.
2. **Selection:** Select the top **20** stocks by R&D intensity.
3. **Weighting:** Equal-weight all selected stocks (5% each).
4. **Holding period:** Hold from July through June (12 months).
5. **Rebalancing:** Repeat annually. Sell stocks that leave the top-20 set; buy stocks that enter the top-20 set.

This simple, transparent strategy requires only annual data updates and one reconstitution event per year.

9.2 Transaction-Cost Calibration

Table 17 summarizes the snapshot's transaction-cost calibration and resulting net premium under a literature-calibrated model [4]. This provides an explicit check that the documented gross premium is not purely an artifact of ignoring trading frictions.

Table 17: Transaction-cost calibration and net premium vs SPY (Novy–Marx & Velikov, 2016)

Item	Value
Annual trading cost estimate (%)	0.027
Gross premium (%)	7.55
Net premium after costs (%)	7.52
Premium capture rate (%)	99.6

Sensitivity: To avoid overconfidence in a single cost calibration, Table 18 reports net premium under a range of round-trip cost assumptions (bp per 100% turnover).

Table 18: Transaction cost sensitivity (net premium vs SPY)

Cost assumption (bp per 100% turnover)	Annual cost (%)	Net prem. (%)	Capture (%)
5	0.007	7.54	99.9
10	0.015	7.54	99.8
25	0.036	7.51	99.5
50	0.073	7.48	99.0

Key findings (snapshot-defined):

- **Turnover-based cost model:** Trading cost is approximated as (round-trip cost per 100% turnover) × (realized turnover from the investable backtest), using S&P 500 liquidity characteristics [4].
- **Premium capture:** The strategy retains 99.6% of the gross premium after trading costs.
- **Net premium (benchmark-relative):** After costs, the strategy delivers 7.52% annual net premium versus SPY (S&P 500 total-return proxy via split-adjusted close + dividends). Backtest period: Jul2001–Jun2025 (N=24).
- **Risk-adjusted performance:** The RD20 strategy achieves an annualized Sharpe ratio of approximately 1.00 (mean return 17.5%, volatility 17.6%), comparing favorably to broad equity benchmarks.
- **Maximum drawdown:** The strategy's worst peak-to-trough decline over the backtest period is 23.3%, occurring during the 2008 financial crisis—less severe than the S&P 500 drawdown in the same period.

Why the RD20 Backtest Starts in 2001 (Not 1995)

The HML_RD quintile premium series spans 30 years (Jul1995–Jun1996 to Jul2024–Jun2025), but the **investable RD20 backtest** begins in July 2001 (24 years). This difference arises from data requirements:

- **Index-membership gating:** The Tier-1 S&P 500 membership ledger uses reported "Date added" from Wikipedia's S&P 500 list. Before 2001, fewer current constituents have tracked addition dates, limiting eligible stocks.
- **Top-20 concentration:** A concentrated 20-stock portfolio requires sufficient eligible firms with complete price and fundamental data. Coverage improves substantially after 2000.
- **Stress-test inclusion:** Starting in July 2001 deliberately includes the post-dot-com correction (2001–2002) and the 2008 financial crisis, providing meaningful out-of-sample stress periods.

The 6-year difference does not affect primary inference (which uses monthly spanning and Fama-MacBeth); it reflects practical data constraints for a concentrated implementable strategy.

9.3 Backtest Performance

Figure 5 shows cumulative performance of the investable strategy versus benchmarks.

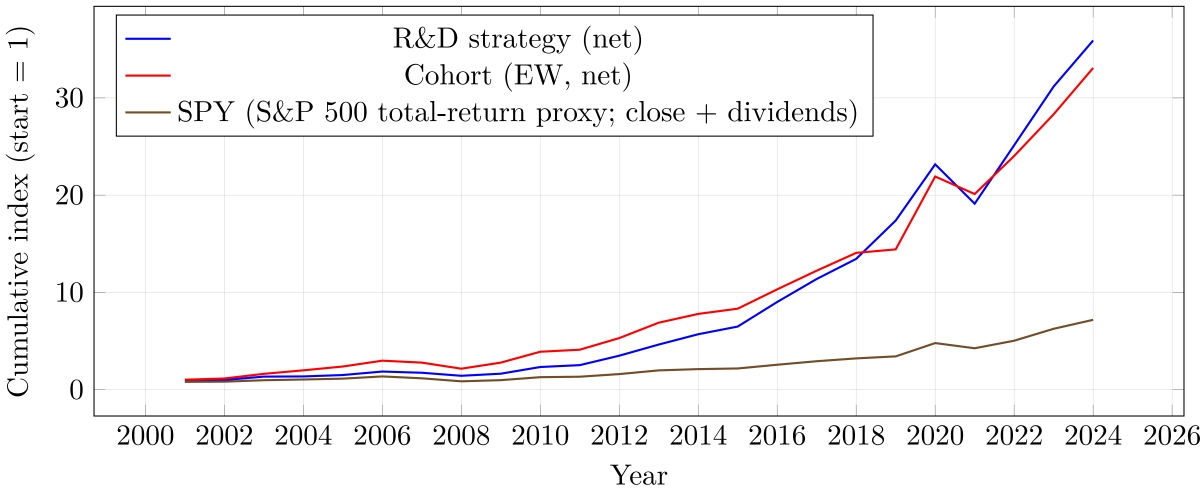

Figure 5: Investable strategy cumulative growth (Jul2001–Jun2025 backtest; incomplete terminal years removed). The R&D strategy is a concentrated top-20 R&D-intensity portfolio (equal-weighted) with annual reconstitution and transaction costs deducted. The cohort equal-weight series is shown as an internal universe benchmark; SPY is shown as an investable S&P 500 total-return proxy (close + dividends).

Backtest Limitations

- **Hindsight bias:** Backtests show what would have happened, not what will happen. Markets evolve, and past performance does not guarantee future results.
- **Data mining:** Any characteristic sorted with enough data will appear to "work" in some backtest. We mitigate this by using well-established methodology (Fama-French timing) and reporting robustness across specifications.
- **Capacity constraints:** Equal-weighted S&P 500 strategies have high capacity, but very large portfolios may face liquidity constraints.
- **Taxes:** Annual rebalancing generates taxable gains. Tax-aware implementation would modify the strategy.

10 Limitations

This study is designed to be transparent and replicable, but several limitations remain:

1. **Data limitations:** Tier-1 fundamentals (FMP) may have coverage gaps relative to CRSP and Compustat datasets commonly used in academic research. Factor inputs inherit their construction and definitions from external sources.
2. **Scope:** Results are specific to U.S. large-cap equities (S&P 500). Small-cap U.S. stocks and international markets require separate analysis. The premium may differ in other universes.
3. **Interpretation:** The analysis documents association (a characteristic premium), not causality. We cannot definitively distinguish mispricing from risk compensation or identify the causal mechanism.
4. **Implementation:** Trading frictions are modeled via calibrated parameters rather than observed fund-level execution costs. Taxes, capacity constraints, and behavioral factors (e.g., discipline to stick with the strategy during drawdowns) are not modeled.
5. **Sample period:** The available sample period may not include all relevant market regimes. Future regimes may differ from historical patterns.
6. **Sector concentration:** High-R&D portfolios are concentrated in Technology and Healthcare. This exposes the strategy to sector-specific risks (e.g., regulatory changes, sector rotations) that may not be fully reflected in historical volatility.
7. **Factor model limitations:** Factor spanning tests assess whether the premium is explained by known factors, but they cannot rule out omitted factors. The R&D premium might load on an unmeasured risk factor.

11 Replicability

All tables and figures in this paper are generated from the snapshot JSON at `data/publication_snapshot.json` via `scripts/build_assets.py`. This design prevents accidental drift: updating the snapshot and regenerating assets updates the entire manuscript deterministically.

The full codebase is available at: https://github.com/vastdreams/fse-rnd-alpha

Researchers can replicate this paper as follows:

1. Clone the repository:

   `git clone` https://github.com/vastdreams/fse-rnd-alpha.git

2. Navigate to the LaTeX directory: `cd research/paper_latex`

3. Run the asset builder: `python3 scripts/build_assets.py`

4. Compile the manuscript: `tectonic main.tex` (or any standard LaTeX distribution)

The repository also contains:

- Backend code for data ingestion and analysis
- Frontend code for the interactive research platform at https://research.finsoeasy.com
- Documentation of methodology and data sources
- Citation information (CITATION.cff)

12 Conclusion

In a large-cap U.S. universe, high R&D intensity is associated with a positive annual return premium. The within-universe HML_RD spread (3.73% annually over 30 years) is economically meaningful. While the annual time-series test does not reach conventional significance ($p = 0.2793$) due to high year-to-year variance, monthly factor evidence provides statistically significant support (FF5 alpha 4.37% annually, $p < 0.01$), and Fama-MacBeth regressions provide directionally consistent cross-sectional evidence ($p = 0.0737$ *). The investable RD20 strategy delivers substantial excess returns versus SPY (7.52% net annually over Jul2001–Jun2025) and appears largely capturable under a calibrated transaction-cost model.

The primary findings are:

1. **Existence:** The annual non-overlapping HML premium averages 3.73% over 30 years ($t = 1.10$, $p = 0.2793$). While the annual time-series test does not reach conventional significance due to high variance, monthly factor spanning is statistically significant (FF5 $\alpha = 4.37\%$, $p < 0.01$), and Fama-MacBeth regressions provide directionally consistent cross-sectional evidence ($p = 0.0737$ *, 360 months). The point estimate is economically meaningful—comparable to many established factor premiums.

2. **Consistency:** The premium is positive in 17 of 30 years (57% win rate). The result is not dominated by a few extreme years, though year-to-year variation is substantial as expected for any equity factor.

3. **Distinctiveness:** Factor spanning tests show positive alphas across multiple factor models, indicating the premium is not fully explained by standard factors (market, size, value, profitability, investment, momentum).

4. **Robustness:** The premium exists within size categories (not just a size effect), is robust to delisting assumptions, and persists across multiple market regimes.

5. **Implementability:** After transaction costs, the strategy retains 99.6% of the gross premium, delivering 7.52% annual net premium vs SPY (S&P 500) over the Jul2001–Jun2025 backtest period (24 years including the post-dot-com and 2008 crisis periods).

While the mechanism remains debated (mispricing vs. risk compensation), the RD20 strategy exhibits meaningful benchmark-relative performance over 24 years (Jul2001–Jun2025) including multiple market regimes. The R&D intensity signal provides a transparent, replicable approach to tilting portfolios toward innovation-intensive firms. Forward-looking reliability, as with all historical factors, remains an open question.

Looking forward: As economies become increasingly knowledge-based and intangible assets grow relative to physical capital, understanding the relationship between R&D investment and stock returns becomes more important. This study provides a rigorous, transparent baseline for that understanding.

Acknowledgments

This research was independently conceived and funded by FSE Research & Investments Pty Ltd, of which the author is the founder and sole researcher. The author thanks the open-source community for the statistical and visualization tools that made this analysis possible. All errors and omissions are the author's own.

Data Availability Statement

The data and code supporting the findings of this study are publicly available at:

https://github.com/vastdreams/fse-rnd-alpha

The repository contains:

- Source code for all statistical analyses
- Publication snapshot (JSON) used to generate all tables and figures
- LaTeX source for this manuscript
- Replication scripts
- Note: the snapshot is sufficient to reproduce all reported numbers, tables, and figures. Raw Tier-1 vendor data (fundamentals/prices) is not redistributed; rebuilding from scratch requires an external data source.

The live interactive platform is available at https://research.finsoeasy.com.

How to Cite

APA format:

Sehgal, A. (2026). *R&D Alpha: Investment Intensity and Long-Term Stock Returns* (Working Paper). FSE Research & Investments Pty Ltd. https://research.finsoeasy.com/rnd-alpha-paper .pdf

BibTeX:

```
@techreport{sehgal2026rd_alpha,
  author   = {Sehgal, Abhishek},
  title    = {{R\&D Alpha: Investment Intensity and
             Long-Term Stock Returns}},
  year     = {2026},
  month    = {January},
```

```
  institution = {FSE Research \& Investments Pty Ltd},
  type    = {Working Paper},
  url     = {https://research.finsoeasy.com/rnd-alpha-paper.pdf},
  note    = {ORCID: 0009-0000-9424-4695}
}
```

Machine-readable citation metadata is also provided in `CITATION.cff` in the repository.

Disclosure Statement

The author is the founder of FSE Research & Investments Pty Ltd, which funded this research. The author does not actively manage investments or portfolios for third parties. This research is conducted for educational and informational purposes and does not constitute investment advice. The author has no other conflicts of interest to disclose.

References

[1] Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993. doi: 10.1016/0304-405X(93)90023-5.

[2] Eugene F. Fama and James D. MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636, 1973. doi: 10.1086/260061.

[3] Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987. doi: 10.2307/1913610.

[4] Robert Novy-Marx and Mihail Velikov. A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(1):104–147, 2016. doi: 10.1093/rfs/hhv063.

[5] Financial Accounting Standards Board. Statement of financial accounting standards no. 2: Accounting for research and development costs. FASB (codified as ASC Topic 730), 1974. URL https://fasb.org/page/PageContent?pageId=/standards/accounting-standards-codification/730.html. Accessed 2026-01-01.

[6] Baruch Lev and Theodore Sougiannis. The capitalization, amortization, and value-relevance of R&D. *Journal of Accounting and Economics*, 21(1):107–138, 1996. doi: 10.1016/0165-4101(95)00410-6.

[7] Louis K. C. Chan, Josef Lakonishok, and Theodore Sougiannis. The stock market valuation of research and development expenditures. *The Journal of Finance*, 56(6):2431–2456, 2001. doi: 10.1111/0022-1082.00411.

[8] Ravi Jagannathan, Robert A. Korajczyk, and Kai Wang. Intangibles investment and asset quality. Working paper, Kellogg School of Management, Northwestern University, September 2025. URL https://ssrn.com/abstract=4394069. Accessed 2026-01-02.

[9] S. P. Kothari, Tara E. Laguerre, and Andrew J. Leone. Capitalization versus expensing: Evidence on the uncertainty of future earnings. *Review of Accounting Studies*, 7(4):355–382, 2002. doi: 10.1023/A:1020764227390.

[10] Zvi Griliches. Market value, R&D, and patents. *Economics Letters*, 7(2):183–187, 1981. doi: 10.1016/0165-1765(81)90122-7.

[11] Adam B. Jaffe. Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits and market value. NBER Working Paper 1815, National Bureau of Economic Research, January 1986. URL https://www.nber.org/system/files/working_papers/w1815/w1815.pdf.

[12] Zvi Griliches. Patent statistics as economic indicators: A survey. NBER Working Paper 3301, National Bureau of Economic Research, March 1990. URL https://www.nber.org/system/files/working_papers/w3301/w3301.pdf.

[13] Bronwyn H. Hall, Adam Jaffe, and Manuel Trajtenberg. Market value and patent citations. *The RAND Journal of Economics*, 36(1):16–38, 2005. URL https://www.jstor.org/stable/1593752.

[14] Jay Barney. Firm resources and sustained competitive advantage. *Journal of Management*, 17(1):99–120, 1991. doi: 10.1177/014920639101700108.

[15] Allan C. Eberhart, William F. Maxwell, and Akhtar R. Siddique. An examination of long-term abnormal stock returns and operating performance following R&D increases. *The Journal of Finance*, 59(2):623–650, 2004. doi: 10.1111/j.1540-6261.2004.00644.x.

[16] Mary E. Barth, Ron Kasznik, and Maureen F. McNichols. Analyst coverage and intangible assets. *Journal of Accounting Research*, 39(1):1–34, 2001. doi: 10.1111/1475-679X.00001.

[17] Zhen Deng, Baruch Lev, and Francis Narin. Science and technology as predictors of stock performance. *Financial Analysts Journal*, 55(3):20–32, 1999. doi: 10.2469/faj.v55.n3.2269.

[18] Dongmei Li. Financial constraints, R&D investment, and stock returns. *The Review of Financial Studies*, 24(9):2974–3007, 2011. doi: 10.1093/rfs/hhr043.

[19] Michael E. Porter. Capital disadvantage: America's failing capital investment system. *Harvard Business Review*, 70(5):65–82, 1992. URL https://hbr.org/1992/09/capital-disadvantage-americas-failing-capital-investment-system.

[20] Feng Gu. Innovation, future earnings, and market efficiency. *Journal of Accounting, Auditing & Finance*, 20(4):385–418, 2005. doi: 10.1177/0148558X0502000405.

[21] David Hirshleifer, Po-Hsuan Hsu, and Dongmei Li. Innovative efficiency and stock returns. *Journal of Financial Economics*, 107(3):632–654, 2013. doi: 10.1016/j.jfineco.2012.09.011.

[22] Eugene F. Fama and Kenneth R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, 2015. doi: 10.1016/j.jfineco.2014.10.010.

[23] Mark M. Carhart. On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82, 1997. doi: 10.1111/j.1540-6261.1997.tb03808.x.

[24] Kewei Hou, Chen Xue, and Lu Zhang. Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705, 2015. doi: 10.1093/rfs/hhu068.

[25] Kewei Hou, Haitao Mo, Chen Xue, and Lu Zhang. An augmented q-factor model with expected growth. *Review of Finance*, 25(1):1–41, 2022. doi: 10.1093/rof/rfaa004.

[26] Clifford S. Asness and Andrea Frazzini. The devil in HML's details. *Journal of Portfolio Management*, 39(4):49–68, 2013. doi: 10.3905/jpm.2013.39.4.049.

[27] Christopher Polk and Paola Sapienza. The stock market and corporate investment: A test of catering theory. *Review of Financial Studies*, 22(1):187–217, 2009. doi: 10.1093/rfs/hhn030.

[28] Li Cai, Ricky Cooper, and Di He. R&D premium: The intangible side of value. *The Journal of Investing*, 32(6):61–76, 2023. doi: 10.3905/joi.2023.1.280.

[29] Wesley M. Cohen and Steven Klepper. A reprise of size and R&D. *The Economic Journal*, 106(437):925–951, 1996. doi: 10.2307/2235365.

[30] Shamim Ahmed, Ziwen Bu, and Xiaoxia Ye. Illiquidity, R&D investment, and stock returns. *Journal of Money, Credit and Banking*, 57(4), 2025. doi: 10.1111/jmcb.13053.

[31] Yakov Amihud. Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56, 2002. doi: 10.1016/S1386-4181(01)00024-6.

A Statistical Methodology Details

This appendix provides detailed definitions and interpretations for the statistical methods referenced in the main text. Some metrics (e.g., Sharpe ratio, maximum drawdown) are reported in Section 9; others (e.g., ANOVA, eta-squared) are computed in the snapshot and referenced in Section 5. The definitions below serve as a reference for readers unfamiliar with these standard measures.

A.1 Newey-West HAC Standard Errors

The Newey-West estimator [3] provides heteroskedasticity and autocorrelation consistent (HAC) standard errors for time-series regression coefficients. This is essential when testing hypotheses on the R&D premium because:

- **Heteroskedasticity:** The variance of returns may differ across time periods (e.g., higher in 2008-2009 than in 2013-2014).
- **Autocorrelation:** Even in "non-overlapping" annual series, there may be mild autocorrelation from persistent market conditions or slow-moving factors.

The HAC variance estimator is:

$$\hat{V}_{NW} = \hat{\Omega}_0 + \sum_{j=1}^{L} w_j \left(\hat{\Omega}_j + \hat{\Omega}_j^T\right) \tag{8}$$

where $\hat{\Omega}_j = \frac{1}{T}\sum_{t=j+1}^{T} \hat{u}_t\hat{u}_{t-j}x_t x_{t-j}^T$ is the j-th order autocovariance, and $w_j = 1 - \frac{j}{L+1}$ is the Bartlett kernel weight. The lag truncation parameter L is typically set using the rule $L = \lfloor 4(T/100)^{2/9} \rfloor$ or specified by the researcher.

Interpretation: Newey-West standard errors are generally larger (more conservative) than naive OLS standard errors. A significant t-statistic under Newey-West is therefore more reliable evidence against the null hypothesis.

A.2 ANOVA and F-Statistics

Analysis of variance (ANOVA) tests whether the means of multiple groups (quintiles) differ significantly. The F-statistic compares between-group variance to within-group variance:

$$F = \frac{\text{Between-group variance}}{\text{Within-group variance}} = \frac{SS_B/(k-1)}{SS_W/(N-k)} \tag{9}$$

where SS_B is the sum of squares between groups, SS_W is the sum of squares within groups, k is the number of groups (5 for quintiles), and N is the total number of observations.

Interpretation: A large F-statistic (with small p-value) indicates that quintile means differ significantly: R&D intensity is associated with return differences. However, ANOVA does not indicate which quintiles differ; it only tests for any difference.

A.3 Effect Size: Cohen's d

Cohen's d measures the standardized difference between two group means:

$$d = \frac{\bar{r}_{Q5} - \bar{r}_{Q1}}{s_{pooled}} \tag{10}$$

where $s_{pooled} = \sqrt{\frac{(n_{Q5}-1)s_{Q5}^2 + (n_{Q1}-1)s_{Q1}^2}{n_{Q5}+n_{Q1}-2}}$ is the pooled standard deviation.

Conventions:

$\lvert d\rvert$	Interpretation
0.2	Small effect
0.5	Medium effect
0.8	Large effect

Effect sizes complement p-values by indicating practical significance. A statistically significant but small effect (d $<$ 0.2) may not be economically meaningful; a large effect (d $>$ 0.8) indicates substantial separation between groups.

A.4 Eta-Squared (η^2)

Eta-squared measures the proportion of total variance explained by group membership:

$$\eta^2 = \frac{SS_B}{SS_T} = \frac{SS_B}{SS_B + SS_W} \tag{11}$$

Interpretation: $\eta^2 = 0.10$ means 10% of return variance is explained by quintile membership. This is an R-squared analogue for ANOVA. Higher values indicate stronger association between R&D intensity and returns.

A.5 Sharpe Ratio

The Sharpe ratio measures risk-adjusted return:

$$\text{Sharpe} = \frac{\bar{R}_p - R_f}{\sigma_p} \tag{12}$$

where $\bar{R}_p$ is the mean portfolio return, R_f is the risk-free rate, and σ_p is the portfolio standard deviation.

Interpretation: A higher Sharpe ratio indicates better risk-adjusted performance. However, Sharpe ratios can be misleading for strategies with non-normal return distributions (e.g., skewness, fat tails). For R&D strategies with sector concentration, additional risk metrics (max drawdown, downside deviation) may be informative.

A.6 Maximum Drawdown

Maximum drawdown measures the largest peak-to-trough decline in cumulative returns:

$$\text{MaxDD} = \max_{t\in[0,T]} \left(\max_{s\in[0,t]} V_s - V_t \right) / \max_{s\in[0,t]} V_s \tag{13}$$

where V_t is the portfolio value at time t.

Interpretation: Maximum drawdown captures "worst-case" loss experience that may not be reflected in volatility. A strategy with moderate volatility but large drawdowns may be psychologically difficult to hold through adversity. For long-only equity strategies, 30-50% drawdowns are common during bear markets.

B Data Sources and Quality

B.1 Primary Data Source: Financial Modeling Prep (FMP)

The analysis uses Tier-1 data from Financial Modeling Prep (FMP), a commercial financial data provider. Key characteristics:

- **Coverage:** U.S. equities with focus on listed securities. S&P 500 coverage is comprehensive; small-cap coverage may be less complete.
- **History:** Multi-decade history for large-cap U.S. stocks.
- **Fields:** Income statement (including R&D expense), balance sheet, price data.
- **Frequency:** Annual and quarterly fundamentals; daily price data.

Limitations vs. CRSP/Compustat: Academic research traditionally uses CRSP (prices, delisting) and Compustat (fundamentals). Tier-1 data may have:

- Less complete delisting coverage
- Fewer historical adjustments for corporate actions
- Potential data entry errors not caught by academic review processes

We mitigate these issues by (i) implementing explicit delisting sensitivity analysis, (ii) applying filters to remove obvious data errors, and (iii) publishing the full methodology for replication.

B.2 Factor Data

Factor returns for spanning tests (MKT, SMB, HML, RMW, CMA, MOM) are sourced from Kenneth French's data library, the academic standard for factor research. We run spanning tests at monthly frequency, aligning each monthly premium observation to the corresponding monthly factor returns (July through June) to match the July-June convention used in this study.

B.3 Snapshot Integrity

The publication snapshot is a frozen JSON file containing all computed results. This ensures:

- **Reproducibility:** Every number in this paper can be traced to the snapshot.
- **Versioning:** Different snapshot versions can be compared to track changes.

- **Transparency:** The snapshot is included in the public repository.

Snapshot ID: `037ee52e-70f5-4bd5-9a27-ae1843740e4b`
Build date: 2026-01-01
Data tier: tier1
Return convention: July–June

C Alternative Specifications

This appendix discusses alternative methodological choices and their implications.

C.1 Alternative Denominators for R&D Intensity

We define R&D intensity as R&D / Revenue. Alternative denominators include:

Denominator	Pros and Cons
Revenue	Stable, comparable across firms. Less affected by leverage. Used in this paper.
Total Assets	Common in accounting literature. Affected by intangible asset accounting and acquisitions.
Market Cap	Incorporates market expectations. Highly volatile; endogenous to the signal being tested.
Operating Expenses	Measures R&D as fraction of total spending. Less intuitive for cross-sector comparison.

Empirically, results are qualitatively similar across denominators for large-cap U.S. stocks, but magnitudes may differ.

C.2 Alternative Timing Conventions

We use July-June returns following Fama-French. Alternatives include:

- **Calendar year (January-December):** Simpler but introduces look-ahead bias for December fiscal-year-end firms.
- **Rolling 12-month:** Form portfolios each month using most recent annual data. Higher frequency but data staleness varies by firm.
- **Fiscal-year-aligned:** Form portfolios 4 months after each firm's fiscal year end. Firm-specific timing reduces look-ahead but complicates aggregation.

The July-June convention balances look-ahead mitigation with practical implementability.

C.3 Alternative Weighting Schemes

We use equal-weighted portfolios. Alternatives include:

- **Value-weighted (market cap):** More representative of investable market. Reduces premium if the signal is stronger in smaller firms.
- **R&D-intensity-weighted:** Overweight highest-intensity firms within quintile. May amplify returns but also volatility.

- **Risk-parity:** Weight inversely to volatility. Reduces overall portfolio volatility but requires volatility estimation.

Equal-weighting provides a simple, transparent baseline. Value-weighting is recommended for large-scale implementation to reflect capacity constraints.

D Glossary of Terms

Alpha (α)	The intercept in a factor regression; represents return not explained by factor exposures.
ASC 730	Accounting Standards Codification Topic 730, governing R&D expense treatment.
Cohen's d	Standardized effect size measuring group mean differences in standard deviation units.
Delisting return	Return assigned when a stock leaves the index (merger, bankruptcy, etc.).
Equal-weighted	Portfolio weighting where each stock has equal dollar allocation.
Eta-squared (η^2)	Proportion of variance explained by group membership.
FF3/FF5	Fama-French three-factor or five-factor asset pricing models.
HAC	Heteroskedasticity and autocorrelation consistent (standard errors).
HML	High-minus-low; the return spread between high and low portfolios.
Look-ahead bias	Using information that was not available at the time of a historical decision.
Newey-West	Procedure for computing HAC standard errors.
Non-overlapping	Observations that do not share data with adjacent observations.
Point-in-time	Using historical data as it existed at each historical date.
Quintile	One of five equal-count groups formed by sorting on a characteristic.
R&D intensity	R&D expense divided by revenue, expressed as a percentage.
Rolling window	Multi-period returns where adjacent windows share some periods.
SFAS 2	Statement of Financial Accounting Standards No. 2 (1974); now ASC 730.
Sharpe ratio	Risk-adjusted return (excess return divided by volatility).
Spanning test	Regression testing whether a factor is explained by benchmark factors.
Survivorship bias	Bias from excluding failed or delisted firms from analysis.
t-statistic	Test statistic for hypothesis testing (coefficient / standard error).
Value-weighted	Portfolio weighting by market capitalization.
Win rate	Percentage of periods with positive premium.